Am. J. Hum. Genet. 56:1487–1492, 1995

Forensic DNA Data Banking by State Crime Laboratories

Jean E. McEwen

Division of Social Science, Ethics, and Law, Eunice Kennedy Shriver Center for Mental Retardation, Waltham, MA; and Boston College Law School, Newton, MA

Summary

This article reports the results of a survey of the responsible crime laboratories in the first 19 states with legislation establishing forensic DNA data banks. The survey inquired into the labs' policies and procedures regarding the collection, storage, and analysis of samples; the retention of samples and data; search protocols; access to samples and data by third parties; and related matters. The research suggests that (1) the number of samples collected from convicted offenders for DNA data banking has far surpassed the number that have been analyzed; (2) data banks have already been used in a small but growing number of cases, to locate suspects and to identify associations between unresolved cases; (3) crime labs currently plan to retain indefinitely the samples collected for their data banks; and (4) the nature and extent of security safeguards that crime labs have implemented for their data banks vary among states. The recently enacted DNA Identification Act (1994) will provide $40 million in federal matching grants to states for DNA analysis activities, so long as states comply with specified quality-assurance standards, submit to external proficiency testing, and limit access to DNA information. Although these additional funds should help to ease some sample backlogs, it remains unclear how labs will allocate the funds, as between analyzing samples for their data banks and testing evidence samples in cases without suspects. The DNA Identification Act provides penalties for the disclosure or obtaining of DNA data held by data banks that participate in CODIS, the FBI's evolving national network of DNA data banks, but individual crime labs must also develop stringent internal safeguards to prevent breaches of data-bank security.

Introduction

An earlier article (McEwen and Reilly 1994) described the first 19 state laws that established forensic DNA data banks—repositories for the long-term storage of DNA collected from specified categories of criminal offenders and of the DNA profiles derived from their analysis. These DNA data banks will increasingly be used by law-enforcement agencies to locate suspects in criminal cases where biological evidence is available. However, their rapid proliferation has also led to some concerns regarding the possibility of their misuse (Scheck 1994).

To gain a better understanding of how forensic DNA data banks will operate, a survey was conducted of the individuals in the responsible crime labs of each state that had a data-banking law as of July 1, 1993. The purpose of the survey, conducted in October 1993, was to learn about the labs' policies and procedures (if any) regarding the collection, storage, and analysis of samples; the retention of samples and data; search protocols; access to samples and data by third parties; and related matters. Because many of the data banks contacted were, at the time of the survey, still in a start-up phase, their policies were not yet clearly defined. As a result, most respondents were unable to provide definitive responses to many questions, making it difficult to draw meaningful conclusions. Nevertheless, some patterns emerged that are interesting, particularly when viewed against the backdrop of the recently enacted DNA Identification Act. The DNA Identification Act, signed into law in September 1994 as part of the comprehensive federal crime legislation, authorizes the FBI director to establish a national DNA identification index. This index, known as CODIS (an acronym for the *C*ombined *D*NA *I*dentification *S*ystem), actually began in 1990 as a pilot program and is expected to be implemented on a national level during 1995. CODIS will enable DNA data banks in different states to exchange investigative leads, so that suspects can be identified and unresolved cases can be linked.

Respondents from 18 of the 19 states contacted for the survey agreed to be interviewed (see table 1); Arizona's data bank, which had begun to collect but not to analyze convicted offender samples, declined to participate but did provide an estimate of the number of samples that that state had collected. Of the 19 labs, 10 were designated pilot sites for the FBI's CODIS program. These labs tended to be somewhat better funded and to be further along in their collection and analysis efforts than the non-CODIS pilot sites; however, no other nota-

Received February 8, 1994; accepted for publication February 16, 1995.

Address for correspondence and reprints: Jean E. McEwen, J.D., Shriver Center for Mental Retardation, Division of Social Science, Ethics, and Law, 200 Trapelo Road, Waltham, MA 02254.

0002-9297/95/5606–0028$02.00

Table 1

Status of DNA Forensic Data-Banking Collection and Analysis Efforts in CODIS Pilot and Nonpilot Laboratories, as of October 1993

State	CODIS Pilot Laboratory	Collection of Samples	Analysis of Samples
Arizona	X	X	
California	X	X	X
Colorado		X	
Florida	X	X	X
Georgia			
Hawaii		X	
Illinois	X	X	X
Iowa			
Kansas	X	X	X
Kentucky		X	
Michigan			
Minnesota	X	X	X
Missouri			
Nevada	X		
Oregon	X	X	X
South Dakota		X	X
Tennessee			
Virginia	X	X	X
Washington	X	X	X

Table 2

Estimated Percentage of Samples Collected from Convicted Offenders Analyzed for DNA Forensic Data Banks, as of October 1993

State	Samples Collected	No. (%) of Samples Analyzed
Arizona	350	0
California	37,000[a]	1,000 (2.7)
Colorado	3,000	0
Florida	5,400	2,900 (53.7)
Georgia	0	NA
Hawaii	100	0
Illinois	5,000	1,800 (36)
Iowa	0	NA
Kansas	3,000	600 (20)
Kentucky	370	0
Michigan	0	NA
Minnesota	3,400	2,500 (73.5)
Missouri	0	NA
Nevada	0	NA
Oregon	4,200	2,200 (52.4)
South Dakota	850[b]	230 (27.1)
Tennessee	0	NA
Virginia	70,000	1,800 (2.6)
Washington	9,200	4,400 (47.8)
Total	141,870	17,430 (12.3)

[a] Includes some samples collected under a preexisting-sex-offender registration statute that authorized the testing of samples by using conventional blood markers.

[b] Includes samples from persons who have been arrested but not yet convicted.

ble differences were observed between the responses of the CODIS pilot labs and the other responses.

The Disparity between Collection and Analysis

Thirteen states had begun to collect samples for their data banks from convicted offenders at the time of the survey (see table 2). The number of samples collected varied widely, with Hawaii having acquired only ~100 samples and California and Virginia—populous states with two of the earliest data-banking laws—having acquired some 37,000 and 70,000 samples, respectively. Cumulatively, in laboratories nationwide, ~141,870 samples had been amassed, with more than three-quarters of those residing in California and Virginia. In Virginia, the large number of samples reflects the state's statutory mandate to collect from *all* convicted felons, including nonviolent offenders; in fact, in that state an estimated 15% of the samples were being provided by white-collar criminals. In California, many of the samples had been collected under a predecessor statute that required registered sex offenders to provide blood samples for conventional serology.

The trend toward expanding the range of offenders required to provide samples for data banking under the relevant state statutes will help to ensure steady growth in the number of samples being held by data banks (McEwen and Reilly 1994). For example, respondents from the two largest data banks, California's and Virginia's, each stated that they would be collecting >10,000 samples annually when their data banks are fully operational and they have caught up with backlog (see table 3). Most states' statutes require taking samples not only from offenders convicted after the statute's effective date but also from those who are already incarcerated. This suggests that, within just a few years, crime labs across the country will collectively have accumulated many hundreds of thousands of DNA samples.

The rate at which DNA samples are being collected,

Table 3

Projected Annual Estimates, for 18 States, of Samples to Be Collected from Convicted Offenders under DNA Data-Banking Laws

	No. of States
≤500	6
501–1,000	2
1,001–5,000	8
5,001–10,000	0
>10,000	2

however, has far outpaced the rate at which they are being analyzed. Primarily because of budgetary constraints, only 9 of the 13 states that were collecting samples had begun to analyze them, and in 2 of those states the testing was being done, either exclusively or in part, by outside labs rather than on site. (Respondents in five other states indicated that they would probably need to contract with private labs to help them analyze some of their data-bank samples in the future.) Collectively, nationwide, only ~17,430 samples had been tested, representing 12.3% of the total samples collected (see table 2). Minnesota had analyzed the largest percentage of its samples; there, nearly three-quarters of all samples from convicted offenders had been analyzed, all of the unanalyzed samples were on membrane, and the lab was reporting a backlog of only 3–4 mo. Washington, which had analyzed some 4,400 of its samples, was the farthest along in terms of raw numbers. On the other hand, California, which in recent years has been plagued by budgetary difficulties, had analyzed <3% of its 37,000 samples. Thus, in that state (as in Virginia, the state that collects from all felons), tens of thousands of samples are awaiting analysis—with more samples coming in every month.

The large disparity between the number of samples collected for data banking and the number of samples analyzed reflects in part both the high cost of setting up a forensic DNA lab and the fact that crime labs have not, at least historically, been well funded for DNA work. Although six respondents indicated that they had in their operating budgets a line item for their DNA data banks, in most states the funds for analyzing data-bank samples were being taken from the lab's overall operating budget for DNA analysis, if any. Eight of the 18 respondents generally described their DNA data-banking budgets for the 1993–94 fiscal year as being either very or somewhat inadequate to meet the expectations imposed by their state's DNA data-banking law. (Interestingly, Louisiana, which previously had a DNA data-banking law, repealed its statute in 1993, because of lack of funding for its implementation.)

Newer, PCR-based identification technologies, as they come on line, should greatly enhance the speed with which samples (currently analyzed by the more time-consuming RFLP technology) can be processed. In addition, the 1994 enactment of the DNA Identification Act should over time should help to ease sample backlogs in at least some states. That act will provide $40 million in federal matching grants, over a 5-year period, to state and local crime labs that are seeking to develop or improve their forensic DNA testing capabilities, so long as the labs comply with the quality-assurance standards issued by the FBI, submit to external proficiency testing, and limit access to DNA information. However, because states will not begin to receive federal funding under the DNA Identification Act until fiscal year 1996, and because much of the funding will be backloaded into subsequent fiscal years, it may be some time before the gap between the number of convicted offender samples collected and the number analyzed will be closed. It also remains unclear how labs will choose to allocate the influx of additional funds and divide up their workloads, as between analyzing samples for their data banks and prioritizing and testing evidence samples in cases without suspects. Historically, labs generally have not analyzed samples in cases without suspects, since, without any reference DNA profiles for comparison, there was little reason for doing so. But until labs have the capacity to analyze in larger numbers *both* their convicted offender samples *and* their unknown evidence samples (including both those in backlogs and those to be acquired in the future), the immediate utility of their DNA data banks is likely to remain limited.

Law-Enforcement Potential and Civil-Liberties Concerns

It appears, however, that where adequate resources exist to analyze samples in both of the above categories, data banks can in fact be a very effective tool in helping to resolve violent-crime cases where no nonbiological leads exist. At the time of the survey in October 1993, Illinois, Minnesota, and Virginia were each reporting successes in achieving "cold hits"—i.e., in identifying suspects in cases where none previously existed—through their data banks, with Minnesota's data bank achieving *two* such matches. The Virginia case involved a rape, the Illinois case involved a murder and attempted murder and sexual assault (two victims), and the Minnesota cases involved one rape and one sexual assault–murder. Interestingly, the match in Virginia occurred on only the sixth attempt to run an unknown evidence profile against the convicted-offender profiles held in the data bank.

DNA data banks can also help to establish *associations* between two or more unresolved cases, such as when banked evidence profiles from two or more unresolved cases are found to match each other. At the time of the survey, both Florida and Minnesota had used their data banks in this manner—and, in the case of Minnesota, several crimes thus linked were later resolved when a suspect was independently identified in a separate investigation and his DNA was found to match that involved in the earlier cases. By October of 1994, 1 year after this survey was conducted, the FBI was reporting 14 "hits" aiding 34 separate investigations, including 7 hits resulting from matches to convicted-offender profiles (United States Department of Justice, Federal Bureau of Investigation 1994).

A data bank may also, on occasion, lead investiga-

tors to the identification of persons whose DNA profiles do not precisely match an evidence profile but that appear to derive from a close biological relative of the perpetrator. The survey did not provide any indication that this specific situation had yet occurred, but few respondents in the survey seemed to have considered this possibility. When asked what procedures they would follow in a case where no precise match occurred but where the search suggested that the evidence sample in question might derive from a relative of someone in the data bank, 12 of the 17 respondents answering this question were unsure what would be done with this information.

Nevertheless, the potential for this to occur may be inherent in the National Academy of Sciences' recommendation that an initial match obtained through a search of a data bank should be confirmed by testing a new sample obtained from the person so identified with additional loci—but that only the statistical frequency associated with the additional loci should be presented as evidence at trial (National Academy of Sciences, National Research Council, Commission on Life Sciences 1992). The purpose of this recommendation is ostensibly to prevent any selection bias inherent in searching a data bank. Given this recommendation, however, labs may be inclined initially to search their data banks by using only two probes—those that are the least polymorphic—in an effort to "save" the most polymorphic probes for confirmation and presentation in evidence at trial. This, in turn, raises the potential for generating multiple suspects after two probes, all or all but one of whose DNA will turn out not to match the evidence sample after further testing but who, if the National Academy of Sciences' directive is interpreted literally, may be unnecessarily subjected to further blood drawing. Even if only one lead is generated based on the two probes, that lead may be to a person (such as a relative of the perpetrator) whose DNA happens to match the evidence sample across two identifying probes but will not match on additional probes. By their nature, cases in which resort must be made to a DNA data bank in order to identify a suspect will tend to be cases in which DNA evidence is more crucial to identity than it is in many other cases, where such evidence may be merely corroborative. As a result, the pressure on law enforcement to go out and draw the blood of potential suspects, with as little "loss of evidence" as possible, may be substantial. Although it is not yet clear how strictly labs will adhere to the language in the National Academy of Sciences report in designing their search protocols, the potential for a data bank to generate whole lists of numerous potential suspects, rather than conclusively pointing to a single individual on four or more probes, will increase in the future, as the number of persons in data banks increases and as labs begin to share DNA data on an interstate basis. This may raise difficult ethical and legal issues.

Three respondents also indicated that their data banks had been used in some cases to *exclude* tentative suspects, in situations where a particular person (such as a habitual sex offender) was initially suspected of committing a crime but was ruled out when his DNA sample *collected for the databank, as a result of a previous conviction,* was tested and found not to match the evidence profile. (These exclusions were in addition to the *non–data bank* exclusions based on DNA evidence, which have now become fairly commonplace—i.e., the more typical case in which a suspect is asked to provide a blood sample for comparison with an evidence sample and in which the two DNA profiles do not match). The finding that DNA data banks have, on occasion, been used for actual exclusions of tentatively identified suspects is important because it suggests that there are instances in which the availability of a data bank may actually spare innocent individuals from having suspicion focused on them and being actively investigated (perhaps to the point of being asked for a confirmatory blood sample), thus indirectly furthering their privacy interests. Although the CODIS system will disallow the amassing of "suspect files," there will be no prohibition against comparing an evidence profile with a DNA profile from a particular convicted offender that is already available in the data bank, where probable cause for a search of that individual exists.

The Retention of Samples

Although in some cases their data-banking legislation does not expressly require it, 11 respondents stated that they planned to destroy samples drawn from convicted offenders, as a matter of course, in cases where the conviction that supplied the basis for drawing the sample was later overturned. On the other hand, several respondents in states where the relevant statutes do not require expungement stated that, once the material was in their possession, they considered it theirs to keep—much as has traditionally been done with conventional fingerprints.

Respondents in all nine states that were already analyzing their samples at the time of the survey were keeping those samples after testing them and stated that, apart from those situations where expungement may be warranted, they planned to do so indefinitely. Although a couple of respondents commented that there would probably come a point when it was logistically impossible to keep all their samples, the consensus was that it was important to keep them in the event that they would be needed at a later date—e.g., to confirm a putative match, without the need to go out and draw another sample from the suspect until the match had been verified on the basis of a retesting of the stored sample.

Several respondents also stated that they wanted to keep samples on hand for retesting at a later date, when newer, PCR-based technologies come on line, so as not to lock themselves permanently into RFLP technology and see their data banks eventually become obsolete. This, however, raises the question of whether the large-scale collection and analysis of samples for data banking at the present time may be premature. On the one hand, it is difficult to argue against a policy of going forward with analyzing at least those samples taken from the most serious habitual offenders; had some states not begun to do this, presumably none of the suspects in the successful searches earlier described would have been identified. On the other hand, any rush to launch a massive data-bank program using RFLP technology may ultimately prove inefficient, since the samples will only need to be reanalyzed when more stable, PCR-based technologies emerge. The National Academy of Sciences, in its 1992 report, specifically cautioned against the rush to develop large RFLP-based data banks, observing that data banks launched too hastily may, in the end, find themselves locked into a "dinosaur" technology (National Academy of Sciences, National Research Council, Commission on Life Sciences 1992). Significantly, too, although that report acknowledged the practical desirability of retaining samples for short periods until the technology becomes more stable, it recommended that, once this has occurred and the data banks are better established, samples should be destroyed "promptly" after they have been analyzed (National Academy of Sciences, National Research Council, Life Sciences Commission 1992).

Access to Banked Samples and Data

The survey's finding that labs plan to retain samples indefinitely after they have analyzed them for their data banks raises obvious questions about data-bank security, because of the potential wealth of highly sensitive, personal genetic-information samples contain. Most of the survey respondents indicated that they planned to implement special confidentiality safeguards for their data banks, such as coding their samples, using freezers with special computer-controlled locks or alarms, and/or assessing penalties against lab personnel for the unauthorized release of samples. However, not all labs had yet put such measures in place, and the large number of employees (20 or 30 in several states) who might at least *potentially* have access to samples increases the risk to sample security. Although the DNA Identification Act provides that a state's right to participate in CODIS will be subject to cancellation if it fails to comply with the act's privacy standards, the sample-security protections afforded by the act may be incomplete. The act mandates the imposition of fines up to $100,000 against persons who, without authorization, obtain DNA samples, but it provides no specific penalties for unauthorized sample *dissemination.* (By contrast, the law does provide penalties for both the unauthorized dissemination and obtaining of DNA *data*).

A particularly difficult issue involves whether anonymous banked samples (i.e., samples from which all individual identifiers have been removed) can or should be made available for genetic research purposes. Most states' data-banking statutes provide little guidance on the use of unidentifiable samples, apart from the general statement that the data banks themselves are to be used for "law enforcement purposes" (McEwen and Reilly 1994); arguably, much genetics research into behavior could be viewed as having at least indirect law-enforcement value (Scheck 1994). Indeed, an Alabama statute enacted in 1994 (after the date of this survey) expressly contemplates the use of anonymous samples collected for Alabama's DNA data bank, even for research that is *not* law-enforcement related; it provides that such samples may be used "to provide data relative to the causation, detection, and prevention of disease or disability" and "[t]o assist in other humanitarian endeavors including, but not limited to, educational research or medical research or development" (Alabama Laws 1994). However, the use, for genetic research, of samples collected from convicted offenders, for forensic DNA data banking, without the informed consent of the subjects may raise significant ethical concerns (Office of Protection from Research Risks 1993).

The DNA *data* derived from the *analysis* of raw samples are less sensitive, from the standpoint of personal privacy, because forensic DNA profiles contain only information relating to individual identification. Nevertheless, this information may be of great interest to third parties interested in tracking the identity of individuals, for reasons unrelated to a specific criminal investigation (e.g., in cases involving questions of child support, paternity, affiliation, or immigration). The FBI, in developing CODIS, is designing computer-security software that will provide enhanced protection for the security of DNA data once it is entered into the system, which will include the encryption of data communications by electronic devices available only to the criminal justice community. However, equally stringent computer-security safeguards should be placed on the computers within the labs that hold any personal identifying information associated with the sample (i.e., information that provides the link between the code number on the sample and the offender's name). Several, but not all, respondents in the survey mentioned that they planned to make those computers subject to a higher level of security.

Although, as noted, the DNA Identification Act now provides penalties for the willful disclosure or obtaining of DNA data held by data banks that are part of the

CODIS network, labs must also put into place adequate safeguards to prevent abuses from occurring in the first place. The General Accounting Office (GAO), in a recent report on the National Crime Information Center (NCIC) network—a network on which the FBI has patterned the CODIS system—found that the NCIC network was vulnerable to misuse from individuals with authorized access (United States General Accounting Office 1993). The GAO listed numerous instances of illegal NCIC file breachings by individuals in 23 states, including both the intentional disclosure of information to private investigators in exchange for money and the alteration or deletion of information in NCIC records. It must be emphasized that, to date, there have been no reported instances of unauthorized dissemination of either DNA samples or DNA data. Nevertheless, in light of the particularly sensitive nature of DNA information, the development of enhanced security safeguards will be crucial to minimize the potential for such abuses.

Conclusion

The growing activity of forensic DNA data banking has received remarkably little public or professional attention—a fact that is striking when one considers that cumulatively, in 19 states across the country, crime labs had by the fall of 1993 already amassed some 142,000 DNA samples from convicted offenders. Since that time, the number of states with forensic DNA data-banking laws has nearly doubled; by October 1994, 31 states had enacted such laws, and bills to establish data banks were pending in several others. The 1994 passage of the DNA Identification Act is likely to provide impetus for the creation of additional DNA data banks and to increase the level of data-banking activity in those states that already have such laws.

The early successes of some of the most active DNA data banks—both in locating suspects in violent crime cases that would otherwise never have been resolved and in identifying associations between groups of unsolved cases—is likely to fuel popular pressure to extend DNA data-banking requirements to a wider range of offenders (McEwen and Reilly 1994). However, state lawmakers should carefully consider the extent of the resources that will be available for the analysis of samples, before they enact broad-based data-banking laws. Currently, the wide scope of coverage under many states' laws has led to the amassing of many more samples than can realistically be processed. Although the expected influx of resources to states, under the DNA Identification Act, should ease some sample backlogs, it may be some time before the gap between collection and analysis is closed so that data banks can realize their full potential.

The large number of samples that have been collected, coupled with the fact that data banks plan to keep their samples indefinitely as a matter of routine after they have been analyzed, makes it essential that ongoing attention be paid to data-bank security. Although the privacy provisions contained in the new DNA Identification Act represent an important step in the right direction, individual crime labs that store samples must also develop strict internal policies to safeguard the privacy of those who provide DNA for their data banks.

Acknowledgments

This work was supported in part by the Human Genome Project, Department of Energy grant DE-FG02-91ER61237, MCH grant MCJ-259151-02-0, ADD grant 90 DD 0213, and Department of Mental Retardation of the Commonwealth of Massachusetts grant 1000-10003-SC. I also acknowledge the support of Boston College Law School.

References

Alabama Laws (1994) 1st Spec Sess Act 94-804 (SB 100)

Ad Hoc Committee on Individual Identification, The American Society of Human Genetics (1990) Individual identification by DNA analysis: points to consider. Am J Hum Genet 46:631-634

DNA Identification Act (1994) PL 103-322, 1994 HR 3355, 108 Stat 1796, §210304

McEwen JE, Reilly PR (1994) A review of state legislation on DNA forensic data banking. Am J Hum Genet 54:941-958

National Academy of Sciences, National Research Council, Commission on Life Sciences (1992) DNA technology in forensic science. National Academy Press, Washington, DC

Office of Protection from Research Risks (1993) Protecting human research subjects: institutional review board guidebook. US Government Printing Office, Washington DC

Scheck B (1994) DNA data banking: a cautionary tale. Am J Hum Genet 54:931-933

United States Department of Justice, Federal Bureau of Investigation (1994) The national DNA identification index (CODIS) program description. Federal Bureau of Investigation, Washington, DC

United States General Accounting Office. 1993. National Crime Information Center: legislation needed to deter misuse of criminal justice information. GAO/T-GGD-93-41. Government Printing Office, Washington, DC